MISY 5366.W01: Data Warehousing and Data Mining for Business Intelligence

Spring 1 Online Mini-Term 20XX

Instructor: Dr. Joseph S. Mollick
Office: OCNR 389
Telephone: 361-825-2853
E-mail: joseph.mollick@tamucc.edu
Office Hours: MWF 8-9a.m & MWF 1:00-2:00p.m.

Required Materials:
Textbook Required
“Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner” by Galit Shmueli, Peter C. Bruce, Nitin R. Patel. Publisher: Wiley; 3rd edition (October 26, 2016)
ISBN-10: 1118729277
Earlier editions will NOT work.

Articles
1. “Big Data, Analytics and the Path From Insights to Value,” by Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins, Nina Kruschwitz HBS SMR372-PDF-ENG.
3. Other articles and or cases will be added to this list by groups of students and the instructor.

Prerequisites:
MISY 5325 Software Based Business Solutions or its equivalent
ORMS 5310 Statistical Decision Analysis or its equivalent

Course Description:
In the information age, organizations can and do collect massive amounts of data. Yet organizations are often "data rich" but "information and knowledge poor". This course is designed to prepare business professionals who, by using analytical methods and data mining and data visualization tools, will be able to harness the potential of data by extracting business intelligence that can be used to improve decisions and operations at various points in the value chain.
Learning Objectives:
By the end of this course, the students will be able to:
1. Define and understand data warehousing and its potential for business intelligence (BBA Goal 3, Objectives 1, 2 and 3; BBA Goal 2, Objectives 1, 2 and 3)
2. Be able to use data mining techniques to extract business intelligence from data (BBA Goal 3, Objectives 1, 2 and 3; BBA Goal 2, Objectives 1, 2 and 3)
3. Be able to create value by analyzing data using data mining and visualization software (BBA Goal 3, Objectives 1, 2 and 3; BBA Goal 2, Objectives 1, 2 and 3; BBA Goal 1, Objectives 1 and 2)
4. Be familiar with major data mining software packages (BBA Goal 1, Objectives 1 and 2)
5. Learn to extract data from various organizational sources such as databases, data warehouses, data marts, data cubes, Excel workbooks and others. (BBA Goal 2, Objectives 1, 2 and 3)
6. Develop data mining applications for various functional areas of an organization (BBA Goal 2, Objectives 1, 2 and 3)
7. Understand the potential impact of business intelligence applications in an organization (BBA Goal 2, Objectives 1, 2 and 3; BBA Goal 4, Objectives 1 and 2)
8. Understand the projected growth of data mining for business intelligence. (BBA Goal 3, Objectives 1, 2 and 3)

Instructional Methodology:
Lecture, group discussions and hands-on use of computers in the classroom.

Relationship to Other Course work:
Data warehousing and data mining can help extract patterns, trends, relationships among variables and insights that can help improve decisions and operations in all functional areas of an organization. To that end this course draws on students’ knowledge from courses in management, marketing, operations, finance, accounting, and human resources management, to build applications that can help an organization achieve its goals. This course directly builds on students' knowledge in business database management, statistical decision analysis and software-based business solutions.

Oral and Written Communication Content:
Class discussions, presentations, discussion forums, quizzes, written reports on statistical analyses of data, research papers as well as major examinations will provide an opportunity for evaluating each student’s performance in oral and written communication.

Technology Applications:
The student is expected to have a good working knowledge of Excel and Access. Basic knowledge of statistics and skills with statistical software tools are expected of each student. This course heavily depends on database technology, electronic spreadsheets and statistical software, and as such it is not for students that are challenged in the area of information technology.
Ethical Perspectives:
Ethical issues and practices of firms, their social responsibility and environmental stewardship will be discussed as they relate to data warehousing, data mining, decision making and management of business organizations.

Global Perspectives:
In today's world, customers, retailers, suppliers and manufacturers are linked globally via computers and communication networks. Data warehousing and data mining solutions will be presented as useful and valuable for all organizations operating locally or globally.

Demographic Diversity Perspectives:
Presentations, discussions, cases, research projects and data mining exercises will show how practices in the area of data warehousing and data mining for business intelligence present opportunities for all individuals and all organizations.

Political, Social, Legal, Regulatory, and Environmental Perspectives:
Examples from other countries will be discussed to compare data warehousing, data mining and business intelligence applications, philosophies and approaches in different countries, societies and legal and regulatory environments. Behavioral and ethical implications of data management procedures and their impact on users will be discussed.

Performance Evaluation and Grading:

| Data Mining Software Discussion (DMSD) | 100 |
| 3 Data Mining Discussions (DMD), 50x3=150 | 150 |
| 1 Data Mining case x 100 | 100 |
| 1 Research paper x 200 | 200 |
| Exam1: MCQ part 100+ Problem part 50 = 150 | 150 |
| Exam2: MCQ part 100+ Problem part 50 = 150 | 150 |
| Comprehensive Final : MCQ part 100+ Problem part 50 = 150 | 150 |
| **Total** | **1000** |

Detailed guideline for each discussion, data mining case and research paper requirements are given in a separate document posted in the content folder in Blackboard.

All course material is considered for exam questions--all assigned readings whether discussed in class or not and all material presented in video lectures and online discussion forums whether covered in assigned readings or not. No make-up exam will be given. A letter grade will be determined based on the total points earned, as follows:
A: A total of 900 or more.  B: A total of 800-899.5.  C: A total of 700-799.5  D: A total of 600-699.5.  F: A total of less than 600.

Late Policy:
Keep track of the due dates for assignments. You may do the assignments before the due date. No excuse for late work will be considered. If not done by the due date, they will automatically receive a grade of zero.
**Attendance Policy:**

In order to achieve the objectives of this course, students are expected to participate in all class activities through discussion forums and other tools. There will be no make-up exams.

**Discussion on Blackboard**

To help students better understand course materials, and to learn from one another in a computer mediated networked environment, the instructor will use discussion forums and other tools in Blackboard.

- If you do not understand a concept or encounter a problem/error that you do not know how to resolve yourself, you will need to post your questions on the discussion board. You are also expected to answer questions asked by your peers. Occasionally, the instructor will also post questions and answers.
- Before posting a question on the discussion forum, it is the student’s responsibility to check and ensure no same question has been posted before.
- By the end of the semester, every student is expected to have asked and answered 10-20 questions in the forum called Data Mining Software Discussion (DMSD). These discussion comments might earn you extra credit toward your course grade. Straight forward questions (for example – how do I get to this transaction, where can I find this button etc.) will not count. Do not post questions just for the sake of posting.

**Steps to Take When You Have Technical Questions:**

When learning any new software package, any learner is bound to have technical questions. Finding answers to these questions is an important part of learning the software. XLMiner's name changed to Analytic Solver Data Mining in 2018. Since the textbook refers to the software as XL Miner but the software that you will use with this book is now called Analytic Solver Data Mining, there are some changes, and this document published at the following URL describes the resulting changes to the book examples and figures.

[https://docs.google.com/document/d/1aG66Q5EN_Cvqml2muUTzTMsM_ifpS1oViR-o4Fbwwo/edit](https://docs.google.com/document/d/1aG66Q5EN_Cvqml2muUTzTMsM_ifpS1oViR-o4Fbwwo/edit)

When you have a question about XL Miner (Analytic Solver Data Mining) that you cannot answer on your own, you should follow the following steps:

1. If you bought a new book, get XL Miner (Analytic Solver Data Mining) by following instructions at [https://www.dataminingbook.com/node/2443](https://www.dataminingbook.com/node/2443)
2. Check the course discussion board, Data Mining Software Discussion, to see if this question has already been resolved by your classmates.
3. If you did not find the solution in Step 1, consult the help library that comes with XL Miner and any online resources and training videos (e.g. youtube) related to XL Miner (Analytic Solver Data Mining):
4. If you did not find the solution in Steps 2 or 3, ask your classmates for help in discussion board named Data Mining Software Discussion.
5. If you did not find the solution in Step 4, ask your instructor or teaching assistant for help.
6. Post the question that you encountered, along with the solution you found, on the course discussion board to get possible extra credit points.

Following the above steps will help you develop the ability to research and answer Data mining software related questions by yourself, and foster collaboration among your peers to find and share
solutions. These skills will serve you well as you continue to use any data analytics after this class is completed.

**Dropping a Class:**

I hope that you never find it necessary to drop this or any other class. However, events can sometimes occur that make dropping a course necessary or wise. Please consult with me before you decide to drop to be sure it is the best thing to do. Should dropping the course be the best course of action, you must initiate the process to drop the course by going to the Student Services Center and filling out a course drop form. Just stopping attendance and participation WILL NOT automatically result in your being dropped from the class. Observe the last day to drop a class with an automatic grade of “W” this term. **March 6** is the last day to drop a class.

**Classroom/Professional Behavior:**

Texas A&M University-Corpus Christi, as an academic community, requires that each individual respect the needs of others to study and learn in a peaceful atmosphere. Under Article III of the Student Code of Conduct, classroom behavior that interferes with either (a) the instructor’s ability to conduct the class or (b) the ability of other students to profit from the instructional program may be considered a breach of the peace and is subject to disciplinary sanction outlined in article VII of the Student Code of Conduct. Students engaging in unacceptable behavior may be instructed to leave the classroom. This prohibition applies to all instructional forums, including classrooms, electronic classrooms, labs, discussion groups, field trips, etc.

**Academic Honesty:**

This course, and all other courses offered by the College of Business (COB), requires all of its students to abide by the COB Student Code of Ethics (available online at [www.cob.tamucc.edu](http://www.cob.tamucc.edu)). Provisions and stipulations in the code are applicable to all students taking College of Business courses regardless of whether or not they are pursuing a degree awarded by the COB.

**Grade Appeals:**

As stated in University Rule 13.02.99.C2, Student Grade Appeals, a student who believes that he or she has not been held to appropriate academic standards as outlined in the class syllabus, equitable evaluation procedures, or appropriate grading, may appeal the final grade given in the course. The burden of proof is upon the student to demonstrate the appropriateness of the appeal. A student with a complaint about a grade is encouraged to first discuss the matter with the instructor. For complete details, including the responsibilities of the parties involved in the process and the number of days allowed for completing the steps in the process, see University Rule 13.02.99.C2, Student Grade Appeals, and University Procedure 13.02.99.C2.01, Student Grade Appeal Procedures. These documents are accessible through the University Rules Web site at [http://www.tamucc.edu/provost/university_rules/index.html](http://www.tamucc.edu/provost/university_rules/index.html). For assistance and/or guidance in the grade appeal process, students may contact the Office of Student Affairs.
Disabilities Accommodations:

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please call or visit Disability Services at (361) 825-5816 in Corpus Christi Hall 116.

If you are a returning veteran and are experiencing cognitive and/or physical access issues in the classroom or on campus, please contact the Disability Services office for assistance at (361) 825-5816.

<table>
<thead>
<tr>
<th>Summary of Topical Coverage:</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining for Business Intelligence (BI) and Big Data</td>
<td>3</td>
</tr>
<tr>
<td>Data Warehousing for Business Intelligence</td>
<td>2</td>
</tr>
<tr>
<td>Virtual Data Warehouse</td>
<td>3</td>
</tr>
<tr>
<td>Data Exploration and Dimension Reduction</td>
<td>1</td>
</tr>
<tr>
<td>Data Visualization (using XLMiner, Spotfire)</td>
<td>4</td>
</tr>
<tr>
<td>Evaluating Classification and Predictive Performance</td>
<td>3</td>
</tr>
<tr>
<td>Prediction and Classification Methods</td>
<td></td>
</tr>
<tr>
<td>Multiple Linear Regression</td>
<td>3</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td>3</td>
</tr>
<tr>
<td>Text Mining</td>
<td>4</td>
</tr>
<tr>
<td>Neural Nets</td>
<td>1</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>3</td>
</tr>
<tr>
<td>Discriminant Analysis</td>
<td>4</td>
</tr>
<tr>
<td>Mining Relationships among Records (association rules and cluster analysis)</td>
<td>4</td>
</tr>
<tr>
<td>Student Projects</td>
<td>3</td>
</tr>
<tr>
<td>Exams</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>45</strong></td>
</tr>
</tbody>
</table>
Day | Tentative Schedule: Chapter_Topic_Page and activities
1 | CHAPTER 1 Introduction 3
   | Data Mining Discussion opens
   | Data Mining Software Discussion opens
2 | CHAPTER 1 Introduction 3
3 | CHAPTER 2 Overview of the Data Mining Process 14
4 | CHAPTER 2 Overview of the Data Mining Process 14
5 | 
6 | 
7 | CHAPTER 3 Data Visualization 50
8 | CHAPTER 3 Data Visualization 50
9 | CHAPTER 5 Evaluating Predictive Performance 101
10 | CHAPTER 5 Evaluating Predictive Performance 101
11 | CHAPTER 5 Evaluating Predictive Performance 101
12 | First comment in Data Mining Discussion 1 due by 11:59PM
13 | 
14 | CHAPTER 6 Multiple Linear Regression 134
15 | CHAPTER 6 Multiple Linear Regression 134
16 | CHAPTER 6 Multiple Linear Regression 134
17 | CHAPTER 6 Multiple Linear Regression 134
18 | CHAPTER 6 Multiple Linear Regression 134
19 | Data Mining Discussion 1 Closes at 11:59PM
20 | 
21 | **Exam 1 Ch 1,2,3,5,6:** between 5AM and 9PM
22 | CHAPTER 7 k-Nearest Neighbors (kNN) 151
23 | CHAPTER 7 k-Nearest Neighbors (kNN) 151
24 | CHAPTER 7 k-Nearest Neighbors (kNN) 151
25 | CHAPTER 10 Logistic Regression 209
26 | First comment in Data Mining Discussion 2 due by 11:59PM
27 | 
28 | CHAPTER 10 Logistic Regression 209
29 | CHAPTER 12 Discriminant Analysis 264
30 | CHAPTER 12 Discriminant Analysis 264
31 | CHAPTER 12 Discriminant Analysis 264
32 | CHAPTER 12 Discriminant Analysis 264
33 | Data Mining Discussion 2 Closes at 11:59PM
34 | 
35 | **Exam 2: Ch 7,10,12** | between 5AM and 9PM
36 | CHAPTER 14 Association Rules and Collaborative Filtering 297
37 | CHAPTER 14 Association Rules and Collaborative Filtering 297
38 | CHAPTER 14 Association Rules and Collaborative Filtering 297
39 | CHAPTER 15 Cluster Analysis 324
40 |
| 41  | 42  | CHAPTER 15 Cluster Analysis 324 |
|     | 43  | CHAPTER 20 Text Mining 436     |
|     | 44  | CHAPTER 20 Text Mining 436     |
|     | 45  | CHAPTER 20 Text Mining 436     |
|     | 46  | First comment in Data Mining Discussion 3 due by 11:59PM |
|     | 47  | Submit Excel file containing worksheets DMSD and MCQ by 11:59PM |
| 48  | 49  | Research Report Due by 11:59PM |
|     | 50  | CHAPTER 21 Cases 454           |
|     | 51  | CHAPTER 21 Cases 454           |
|     | 52  | CHAPTER 21 Cases 454           |
|     | 53  | Submit Data Mining Case Report by 5PM |
|     | 54  | Data Mining Discussion 3 Closes at 11:59PM |
|     | 55  |                                 |
| 56  | 57  | Study for Final Exam           |
|     |     | **Comprehensive Final Exam**   |
|     |     | between 5AM and 9PM, End of Semester |